

# Min-Max Graph Partitioning and Small Set Expansion

Nikhil Bansal  
IBM Research

Uriel Feige  
Weizmann Institute of Science

Robert Krauthgamer  
Weizmann Institute of Science

Konstantin Makarychev  
IBM Research

Viswanath Nagarajan  
IBM Research

Joseph (Seffi) Naor  
Technion

Roy Schwartz  
Technion

**Abstract**— We study graph partitioning problems from a *min-max* perspective, in which an input graph on  $n$  vertices should be partitioned into  $k$  parts, and the objective is to minimize the maximum number of edges leaving a single part. The two main versions we consider are: (i) the  $k$  parts need to be of equal size, and (ii) the parts must separate a set of  $k$  given terminals. We consider a common generalization of these two problems, and design for it an  $O(\sqrt{\log n \log k})$ -approximation algorithm. This improves over an  $O(\log^2 n)$  approximation for the second version due to Svitkina and Tardos [22], and roughly  $O(k \log n)$  approximation for the first version that follows from other previous work. We also give an improved  $O(1)$ -approximation algorithm for graphs that exclude any fixed minor.

Our algorithm uses a new procedure for solving the Small-Set Expansion problem. In this problem, we are given a graph  $G$  and the goal is to find a non-empty set  $S \subseteq V$  of size  $|S| \leq \rho n$  with minimum edge-expansion. We give an  $O(\sqrt{\log n \log(1/\rho)})$  bicriteria approximation algorithm for the general case of Small-Set Expansion, and  $O(1)$  approximation algorithm for graphs that exclude any fixed minor.

## 1. INTRODUCTION

We study graph partitioning problems from a *min-max* perspective. Typically, graph partitioning problems ask for a partitioning of the vertex set of an undirected graph, under some problem-specific constraints on the different parts, e.g., balanced partitioning or separating terminals, and the objective is *min-sum*, i.e., minimizing the total weight of the edges connecting different parts. In the min-max variant of these problems, the goal is different — minimize the weight of the edges leaving a single part, taking the maximum over the different parts. A canonical example, that we consider throughout the paper, is the Min-Max  $k$ -Partitioning problem: given an undirected graph  $G = (V, E)$  with nonnegative edge weights and  $k \geq 2$ , partition the vertices into  $k$  (roughly) equal parts  $S_1, \dots, S_k$  so as to minimize  $\max_i \delta(S_i)$ , where  $\delta(S)$  denotes the sum of the

edge weights in the cut  $(S, V \setminus S)$ . We design a bicriteria approximation algorithm for this problem. Throughout, let  $w : E \rightarrow \mathbb{R}^+$  denote the edge-weights and let  $n = |V|$ .

Min-max partitions arise naturally in many settings. Consider the following application in the context of cloud computing, which is a special case of the general graph-mapping problem considered in [4] (and also implicit in other previous works [23], [24], [8]). There are  $n$  processes communicating with each other, and there are  $k$  machines, each having a bandwidth capacity  $C$ . The goal is to allocate the processes to machines in a way that balances the load (roughly  $n/k$  processes per machine), and meets the outgoing bandwidth requirement. Viewing the processes as vertices and the traffic between them as edge weights, we get the Min-Max  $k$ -Partitioning problem. In general, balanced partitioning (either min-sum or min-max) is at the heart of many heuristics used in a wide range of applications, including VLSI layout, circuit testing and simulation, parallel scientific processing, and sparse linear systems.

Balanced partitioning, particularly the min-sum version, has been studied extensively during the last two decades, with impressive results and connections to several fields of mathematics, see e.g. [16], [9], [17], [3], [2], [13], [14], [6]. The min-max variants, in contrast, have received much less attention. Previously, no approximation algorithm for the Min-Max  $k$ -Partitioning problem was given explicitly, and the approximation that follows from known results is not smaller than  $O(k\sqrt{\log n})$ .<sup>1</sup> We improve this dependence on  $k$  significantly.

An important tool in our result above is an approximation algorithm for the Small-Set Expansion (SSE) problem. This problem was suggested recently by Raghavendra and Steurer [19] (see also [20], [21]) in the context of the unique games conjecture. Recall that the *edge-expansion* of a subset

Uriel Feige is supported in part by the Israel Science Foundation (grant #873/08).

Robert Krauthgamer is supported in part by The Israel Science Foundation (grant #452/08), and by a Minerva grant.

Joseph (Seffi) Naor is supported in part by the Israel Science Foundation (grant #954/11), and the US-Israel Binational Science Foundation (grant #2010426).

<sup>1</sup>One could reduce the problem to the min-sum version of  $k$ -partitioning. The latter admits bicriteria approximation  $O(\sqrt{\log n \log k})$  [13], but the reduction loses another factor of  $k/2$ . Another possibility is to repeatedly remove  $n/k$  vertices from the graph, paying again a factor of  $k/2$  on top of the approximation in a single iteration, which is, say,  $O(\log n)$  by [18].

$S \subseteq V$  with  $0 < |S| \leq \frac{1}{2}|V|$  is

$$\Phi(S) := \frac{\delta(S)}{|S|}.$$

The input to the SSE problem is an edge-weighted graph and  $\rho \in (0, \frac{1}{2}]$ , and the goal is to compute

$$\Phi_\rho := \min_{|S| \leq \rho n} \Phi(S).$$

Raghavendra, Steurer and Tetali [20] designed for SSE an algorithm that approximates the expansion within  $O(\sqrt{(1/\Phi_\rho) \log(1/\rho)})$  factor of the optimum, while violating the bound on  $|S|$  by no more than a constant factor (namely, a bicriteria approximation). Notice that the approximation factor depends on  $\Phi_\rho$ ; this is not an issue if every small set expands well, but in general  $\Phi_\rho$  can be as small as  $1/\text{poly}(n)$ , in which case this guarantee is quite weak.

One can achieve a true approximation of  $O(\log n)$  for SSE using [18], for any value of  $\rho$ .<sup>2</sup> If one desires a better approximation, then an approximation of  $O(\sqrt{\log n})$  using [2] can be achieved at the price of slightly violating the size constraint, namely a bicriteria approximation algorithm. However, unlike the former which works for any value of  $\rho$ , the latter works only for  $\rho = \Omega(1)$ . In our context of min-max problems we need the case  $\rho = 1/k$ , where  $k = k(n)$  is part of the input. Therefore, it is desirable to extend the  $O(\sqrt{\log n})$  bound of [2] to a large range of values for  $\rho$ .

### 1.1. Main Results

Our two main results are bicriteria approximation algorithms for the Min-Max  $k$ -Partitioning and SSE problems, presented below. The notation  $O_\varepsilon(t)$  hides multiplicative factors depending on  $\varepsilon$ , i.e., stands for  $O(f(\varepsilon) \cdot t)$ .

**Theorem 1.** *For every positive constant  $\varepsilon > 0$ , Min-Max  $k$ -Partitioning admits a bicriteria approximation of  $(O_\varepsilon(\sqrt{\log n \log k}), 2 + \varepsilon)$ .*

This theorem provides a polynomial-time algorithm that with high probability outputs a partition  $S_1, \dots, S_k$  such that  $\max_i |S_i| \leq (2 + \varepsilon) \frac{n}{k}$  and  $\max_i \delta(S_i) \leq O(\sqrt{\log n \log k}) \text{OPT}$ , where OPT is the optimal min-max value of partitioning into  $k$  equal-size parts. (The guarantee on part size can be improved slightly to  $2 - \frac{1}{k} + \varepsilon$ ). This result is most interesting in the regime  $1 \ll k \ll n$ .

**Theorem 2.** *For every positive constant  $\varepsilon > 0$ , Small-Set Expansion admits a bicriteria approximation of  $(O_\varepsilon(\sqrt{\log n \log(1/\rho)}), 1 + \varepsilon)$ .*

This theorem provides a polynomial-time algorithm that with high probability outputs a set  $S$  of size  $0 < |S| \leq (1 + \varepsilon)\rho n$  whose edge-expansion is  $\delta(S)/|S| =$

<sup>2</sup>For very small values of  $\rho$ , roughly  $\rho n \leq O(\log^2 n)$ , a better approximation ratio is known [10].

$O(\sqrt{\log n \log(1/\rho)}) \text{OPT}$ , where OPT is the minimum edge-expansion over all sets of size at most  $\rho n$ . Our algorithm actually handles a more general version, called Weighted Small-Set Expansion, which is required in Theorem 1. We defer the precise details to Section 2.

### 1.2. Additional Results and Extensions

*$\rho$ -Unbalanced Cut:* Closely related to the SSE problem is the following  $\rho$ -Unbalanced Cut problem: The input is again a graph  $G = (V, E)$  with nonnegative edge-weights and a parameter  $\rho \in (0, \frac{1}{2}]$ , and the goal is to find a subset  $S \subseteq V$  of size  $|S| = \rho n$  that minimizes  $\delta(S)$ . The relationship between this problem and SSE is similar to the one between Balanced Cut and Sparsest Cut, and thus Theorem 2 yields the following result.

**Theorem 3.** *For every constant  $0 < \varepsilon < 1$ , the  $\rho$ -Unbalanced Cut problem admits a bicriteria approximation of  $(O_\varepsilon(\sqrt{\log n \log(1/\rho)}), \Omega(1), 1 + \varepsilon)$ .*

This theorem says that there is a polynomial-time algorithm that with high probability finds  $S \subseteq V$  of size  $\Omega(\rho n) \leq |S| \leq (1 + \varepsilon)\rho n$  and value  $\delta(S) \leq O_\varepsilon(\sqrt{\log n \log(1/\rho)}) \text{OPT}$ , where OPT is the value of an optimal solution to  $\rho$ -Unbalanced Cut. This result generalizes the bound of [2] from  $\rho = \Omega(1)$  to any value of  $\rho \in (0, \frac{1}{2}]$ . Our factor is better than the  $O(\log n)$  true approximation ratio that follows from [18], at the price of slightly violating the size constraint. Our algorithm actually handles a more general version, called Weighted  $\rho$ -Unbalanced Cut, which is required in Theorem 1. We defer the precise details to Section 2.4.

*Min-Max-Multiway-Cut:* We also consider the following Min-Max-Multiway-Cut problem, suggested by Svitkina and Tardos [22]: the input is an undirected graph with nonnegative edge-weights and  $k$  terminal vertices  $t_1, \dots, t_k$ , the goal is to partition the vertices into  $k$  parts  $S_1, \dots, S_k$  (not necessarily balanced), under the constraint that each part contains exactly one terminal, so as to minimize  $\max_i \delta(S_i)$ . They designed an  $O(\alpha \log n)$ -approximation algorithm for this problem, where  $\alpha$  is the approximation factor known for Minimum Bisection. Plugging  $\alpha = O(\log n)$ , due to Racke [18], the algorithm of Svitkina and Tardos achieves  $O(\log^2 n)$ -approximation. Using a similar algorithm to the one in Theorem 1, we obtain a better approximation factor.

**Theorem 4.** *Min-Max-Multiway-Cut admits an  $O(\sqrt{\log n \log k})$ -approximation algorithm.*

Somewhat surprisingly, we show that removing the dependence on  $n$  for Min-Max-Multiway-Cut (even though no balance is required) appears hard, which stands in contrast to its min-sum version, known as Multiway Cut, which admits  $O(1)$ -approximation [5], [11]. The idea is to show that it would imply a similar independence of  $n$  for the min-sum version of  $k$ -partitioning, thus for large but constant  $k$ ,

we would get an  $(O(1), O(1))$ -bicriteria approximation for Min-Sum  $k$ -Partitioning, which seems unlikely based on the current state of art [2], [1], [13].

**Theorem 5.** *If there is a  $k^{1-\varepsilon}$ -approximation algorithm for Min-Max-Multiway-Cut for some constant  $\varepsilon > 0$ , then there is a  $(k^2, \gamma)$  bicriteria approximation algorithm for Min-Sum  $k$ -Partitioning with  $\gamma \leq 3^{2/\varepsilon}$ .*

The proofs of Theorems 4 and 5 appear in the full version of the paper.

*Excluded-minor graphs:* Finally, we obtain an improved approximation – *constant factor* – for SSE in graphs excluding a fixed minor.

**Theorem 6.** *For every constant  $\varepsilon > 0$ , Small-Set Expansion admits:*

- *bicriteria approximation of  $(O_\varepsilon(r^2), 1+\varepsilon)$  on graphs excluding a  $K_{r,r}$ -minor.*
- *bicriteria approximation of  $(O_\varepsilon(\log g), 1+\varepsilon)$  on graphs of genus  $g \geq 1$ .*

These bounds extend to the  $\rho$ -Unbalanced Cut problem, and by plugging them into the proof of Theorems 1 and 4, we achieve an improved approximation ratio of  $O(r^2)$  for Min-Max  $k$ -Partitioning and Min-Max-Multiway-Cut in graphs excluding a  $K_{r,r}$ -minor.

### 1.3. Techniques

For clarity, we restrict the discussion here mostly to our main application, Min-Max  $k$ -Partitioning. Our approach has two main ingredients. First, we reduce the problem to a weighted version of SSE, showing that an  $\alpha$  (bicriteria) approximation for the latter can be used to achieve  $O(\alpha)$  (bicriteria) approximation for Min-Max  $k$ -Partitioning. Second, we design an  $O_\varepsilon(\sqrt{\log n \log(1/\rho)})$  (bicriteria) approximation for weighted SSE (recall that in our applications  $\rho = 1/k$ ).

Let us first examine SSE, and assume for simplicity of presentation that  $\rho = 1/k$ . Note that SSE bears obvious similarity to both Balanced Cut and min-sum  $k$ -partition (its solution contains a single cut with a size condition, as in Balanced Cut, but the size of this cut is  $n/k$  similarly to the  $k$  pieces in min-sum  $k$ -partition). Thus, our algorithm is inspired by, but different from, the approximation algorithms known for these two problems [2], [13]. As in these two problems, we use a semidefinite programming (SDP) relaxation to compute an  $\ell_2^2$  metric on the graph vertices. However, new spreading constraints are needed since SSE is highly asymmetric in its nature — it contains only a *single* cut of size  $n/k$ . We devise a randomized rounding procedure based on the *orthogonal separator* mechanics, first introduced by Chlamtac, Makarychev, and Makarychev [7] in the context of unique games. These ideas lead to an algorithm that computes a cut  $S$  of *expected size*  $|S| \leq O(n/k)$  and of *expected cost*  $\delta(S) \leq O(\sqrt{\log n \log k})$  times the SDP

value. An obvious concern is that both properties occur in only expectation and might be badly correlated, e.g., the expected edge-expansion  $\mathbb{E}[\delta(S)/|S|]$  might be extremely large. Nevertheless, we prove that with good probability,  $|S| = O(n/k)$  and  $\delta(S)/|S|$  is sufficiently small.

For SSE on excluded-minor and bounded-genus graphs, we give a better approximation guarantees, of a constant factor, by extending the notion of orthogonal separators to linear programs (LPs) and designing such low-distortion “LP separators” for these special graph families. The proof uses the probabilistic decompositions of Klein, Plotkin, and Rao [12] and Lee and Sidiropoulos [15]. We believe that this result may be of independent interest. Let us note that the LP formulation for SSE is not trivial and requires novel spreading constraints. We remark that even on planar graphs, the decomposition of Räcke [18] suffers an  $\Omega(\log n)$  loss in the approximation guarantee, and thus does not yield  $o(\log n)$  ratio for SSE on this class of graphs.

Several natural approaches for designing an approximation algorithm for Min-Max  $k$ -Partitioning fail. First, reducing the problem to trees à la Räcke [18] is not very effective, because there might not be a single tree in the distribution that preserves *all* the  $k$  cuts simultaneously. Standard arguments show that the loss might be a factor of  $O(k \log n)$  in the case of  $k$  different cuts. Second, one can try and formulate a relaxation for the problem. However, the natural linear and semidefinite relaxations both have large integrality gaps. As a case study, consider for a moment Min-Max-Multiway-Cut. The standard linear relaxation of Calinescu, Karloff and Rabani [5] was shown by Svitkina and Tardos [22] to have an integrality gap of  $k/2$ . In the full version of the paper, we extend this gap to the semidefinite relaxation that includes all  $\ell_2^2$  triangle inequality constraints. A third attempt is to repeatedly remove from the graph, using SSE, pieces of size  $\Theta(n/k)$ . However, by removing the “wrong” vertices from the graph, this process might arrive at a subgraph where every cut of  $\Theta(n/k)$  vertices has edge-weight greater by a factor of  $\Theta(k)$  than the original optimum (see the full version of the paper for details). Thus, a different approach is needed.

Our approach is to use multiplicative weight-updates on top of the algorithm for weighted SSE. This yields a collection  $\mathcal{S}$  of sets  $S$ , all of size  $|S| = \Theta(n/k)$  and cost  $\delta(S) \leq O(\sqrt{\log n \log k})\text{OPT}$ , that covers every vertex  $v \in V$  at least  $\Omega(n/k)$  times. (Alternatively, this collection  $\mathcal{S}$  can be viewed as a fractional solution to a configuration LP of exponential size.) Next, we randomly sample sets  $S_1, \dots, S_t$  from  $\mathcal{S}$  till  $V$  is covered, and derive a partition given by  $P_1 = S_1$ ,  $P_2 = S_2 \setminus S_1$ , and in general  $P_i = S_i \setminus (\cup_{j < i} S_j)$ . This step is somewhat counter-intuitive, since the sets  $P_i$  may have very large cost  $\delta(P_i)$  (because a set  $P_i$  might be a strict subset of a set  $S_{i'}$ ). We show that the total expected boundary of the partition is not very large, i.e.,  $\mathbb{E}[\sum_i \delta(P_i)] \leq O(k\sqrt{\log n \log k})\text{OPT}$ . Then, we start

fixing the partition by the following local operation: find a  $P_i$  violating the constraint  $\delta(P_i) \leq O(\sqrt{\log n \log k})\text{OPT}$ , replace it with the unique  $S_i$  containing it, and adjust other sets  $P_j$  accordingly. Somewhat surprisingly, we prove that this local fixing procedure terminates (quickly). Finally, the resulting partition consists of sets  $P_i$ , each of which satisfies the necessary properties, but now the number of these sets might be very large. So the last step is to merge small sets together. We show that this can be done while maintaining simultaneously the constraints on the sizes and on the costs of the sets.

*Organization:* We first show in Section 2 how to approximate Weighted Small-Set Expansion (in both general and excluded-minor graphs). We then prove in Section 2.4 that Weighted Small-Set Expansion and Weighted  $\rho$ -Unbalanced Cut are equivalent in terms of approximation, up to constant factors. In Section 3 we present an approximation algorithm for Min-Max  $k$ -Partitioning that uses the aforementioned algorithm for  $\rho$ -Unbalanced Cut (and in turn the one for Weighted Small-Set Expansion). Due to space constraints, we defer many details to the full version of the paper.

## 2. APPROXIMATION ALGORITHMS FOR SMALL SET EXPANSION

In this section we design approximation algorithms for the Small-Set Expansion problem. Our main result is for general graphs and uses an SDP relaxation. It actually holds for a slight generalization of the problem, where expansion is measured with respect to vertex weights (see Definition 7 and Theorem 8). We further obtain improved approximation for certain graph families such as planar graphs (see Section 2.3).

To simplify notation, we shall assume that vertex weights are normalized: we consider measures  $\mu$  and  $\eta$  with  $\mu(V) = \eta(V) = 1$ . We denote  $\mu(u) = \mu(\{u\})$  and  $\eta(u) = \eta(\{u\})$ . We let  $(V, w)$  denote a complete (undirected) graph on vertex set  $V$  with edge-weight  $w(u, v) = w(v, u) \geq 0$  for every  $u \neq v \in V$ . In our context, such  $(V, w)$  can easily model a specific edge set  $E$ , by simply setting  $w(u, v) = 0$  for every non-edge  $(u, v) \notin E$ . Recall that we let  $\delta(S) := \sum_{u \in S, v \in V \setminus S} w(u, v)$  be the total weight of edges crossing the cut  $(S, V \setminus S)$ , and further let  $w(E)$  denote the total weight of all edges.

**Definition 7** (Weighted Small-Set Expansion). *Let  $G = (V, w)$  be a graph with nonnegative edge-weights, and let  $\mu$  and  $\eta$  be two measures on the vertex set  $V$  with  $\mu(V) = \eta(V) = 1$ . The weighted small set expansion with respect to  $\rho \in (0, 1/2]$  is*

$$\Phi_{\rho, \mu, \eta}(G) = \min \left\{ \frac{\delta(S)}{w(E)} \times \frac{1}{\eta(S)} : \eta(S) > 0, \mu(S) \leq \rho \right\}.$$

**Theorem 8** (Approximating SSE). *(I) For every fixed  $\varepsilon > 0$ , there is a polynomial-time algorithm that given as input an*

*edge-weighted graph  $G = (V, w)$ , two measures  $\mu$  and  $\eta$  on  $V$  ( $\mu(V) = \eta(V) = 1$ ), and some  $\rho \in (0, 1/2]$ , finds a set  $S \subset V$  satisfying  $\eta(S) > 0$ ,  $\mu(S) \leq (1 + \varepsilon)\rho$  and*

$$\frac{\delta(S)}{w(E)} \times \frac{1}{\eta(S)} \leq D \times \Phi_{\rho, \mu, \eta}(G), \quad (1)$$

where  $D = O_\varepsilon(\sqrt{\log n \log(1/\rho)})$ .

*(II) When the input contains in addition a parameter  $H \in (0, 1)$ , the algorithm finds a non-empty set  $S \subset V$  satisfying  $\mu(S) \leq (1 + \varepsilon)\rho$ ,  $\eta(S) \in [\Omega(H), 2(1 + \varepsilon)H]$ , and*

$$\frac{\delta(S)}{w(E)} \cdot \frac{1}{\eta(S)} \leq D \min_{\substack{\eta(S) \in [H, 2H] \\ \mu(S) \leq \rho}} \left\{ \frac{\delta(S)}{w(E)} \times \frac{1}{\eta(S)} \right\}, \quad (2)$$

where  $D = O_\varepsilon(\sqrt{\log n \log(\max\{1/\rho, 1/H\})})$ .

We prove part I of the theorem in Section 2.1, and part II in Section 2.2. These algorithms require the following notion of  $m$ -orthogonal separators due to Chlamtac, Makarychev, and Makarychev [7].

**Definition 9** (Orthogonal Separators). *Let  $X$  be an  $\ell_2^2$  space (i.e., a collection of vectors satisfying  $\ell_2^2$  triangle inequalities). We say that a distribution over subsets of  $X$  is an  $m$ -orthogonal separator of  $X$  with distortion  $D$ , probability scale  $\alpha > 0$  and separation threshold  $\beta < 1$  if the following conditions hold for  $S \subset X$  chosen according to this distribution:*

- For all  $u \in X$  we have  $\Pr(u \in S) = \alpha \|u\|^2$ .
- For all  $u, v \in X$  with  $\|u - v\|^2 \geq \beta \min(\|u\|^2, \|v\|^2)$ ,

$$\Pr(u \in S \text{ and } v \in S) \leq \frac{\min\{\Pr(u \in S), \Pr(v \in S)\}}{m}.$$

- For all  $u, v \in X$  we have

$$\Pr(I_S(u) \neq I_S(v)) \leq \alpha D \times \|u - v\|^2,$$

where  $I_S$  is the indicator function of the set  $S$ .

**Theorem 10** ([7]). *There exists a polynomial-time randomized algorithm that given a set of vectors  $X$ , positive number  $m$ , and  $\beta < 1$  generates  $m$ -orthogonal separator with distortion  $D = O_\beta(\sqrt{\log |X| \log m})$  and scale  $\alpha \geq 1/p(|X|)$  for some polynomial  $p$ .*

In the original paper [7], the second requirement in the definition of orthogonal separators was slightly different, however, exactly the same algorithm and proof works in our case: If  $\|u - v\|^2 \geq \beta \|u\|^2$  and  $\|u\|^2 \leq \|v\|^2$ , then  $\langle u, v \rangle = (\|u\|^2 + \|v\|^2 - \|u - v\|^2)/2 \leq ((1 - \beta)\|u\|^2 + \|v\|^2)/2 \leq (1 - \beta/2)\|v\|^2$ . Then, by Lemma 4.1 in [7],  $\langle \varphi(u), \varphi(v) \rangle \leq (1 - \beta/2)$ ; hence  $\|\varphi(u) - \varphi(v)\|^2 \geq \beta > 0$  and, in Corollary 4.6,  $\|\psi(u) - \psi(v)\| \geq 2\gamma = \sqrt{\beta}/4 > 0$ .

## 2.1. Algorithm I: Small-Set Expansion in General Graphs

We now prove part I of Theorem 8.

*SDP Relaxation:* In our relaxation we introduce a vector  $\bar{v}$  for every vertex  $v \in V$ . In the intended solution of the SDP corresponding to the optimal solution  $S \subset V$ ,  $\bar{v} = 1$  (or, a fixed unit vector  $e$ ), if  $v \in S$ ; and  $\bar{v} = 0$ , otherwise. The objective is to minimize the fraction of cut edges

$$\min \frac{1}{w(E)} \sum_{(u,v) \in E} w(u,v) \|\bar{u} - \bar{v}\|^2.$$

We could constrain all vectors  $\bar{v}$  to have length at most 1, i.e.  $\|\bar{v}\|^2 \leq 1$ , but it turns out our algorithm never uses this constraint. We require that the vectors  $\{\bar{v} : v \in V\} \cup \{0\}$  satisfy  $\ell_2^2$  triangle inequalities i.e., for every  $u, v, w \in V$ ,  $\|\bar{u} - \bar{w}\|^2 \leq \|\bar{u} - \bar{v}\|^2 + \|\bar{v} - \bar{w}\|^2$ ,  $\|\bar{u}\|^2 \leq \|\bar{u} - \bar{v}\|^2 + \|\bar{v}\|^2$ ,  $\|\bar{u} - \bar{w}\|^2 \leq \|\bar{u}\|^2 + \|\bar{w}\|^2$ . Suppose now that we have approximately guessed the measure  $H$  of the optimal solution  $H \leq \eta(S) \leq 2H$  (this step is not necessary but it simplifies the exposition; in fact, we could simply let  $H = 1$ , since the SDP is otherwise homogeneous). This can be done since the measure of every set  $S$  lies in the range from  $\eta(u)$  to  $n\eta(u)$ , where  $u$  is the heaviest element in  $S$ , hence  $H$  can be chosen from the set  $\{2^t \eta(u) : u \in V, t = 0, \dots, \lfloor \log_2 n \rfloor\}$  of size  $O(n \log n)$ . Then we add a constraint

$$\sum_{v \in V} \|\bar{v}\|^2 \eta(v) \geq H. \quad (3)$$

Finally, we introduce new spreading constraints: for every  $u \in V$ ,

$$\sum_{v \in V} \mu(v) \cdot \min\{\|\bar{u} - \bar{v}\|^2, \|\bar{u}\|^2\} \geq (1 - \rho) \|\bar{u}\|^2.$$

(Alternatively, we could use a slightly simpler, almost equivalent constraint  $\sum_{v \in V} \langle \bar{u}, \bar{v} \rangle \mu(v) \leq \rho \|\bar{u}\|^2$ . We chose to use the former formulation because an analogous constraint can be written in a linear program, see Section 2.3.) In the intended solution this constraint is satisfied, since if  $u \in S$ , then  $\bar{u} = 1$  and the sum above equals  $\mu(V \setminus S) \geq 1 - \rho$ . If  $u \notin S$ , then  $\bar{u} = 0$  and both sides of the constraint equal 0.

The SDP relaxation used in our algorithm is presented below in its entirety. Note that the second constraint can be written as  $\langle \bar{u}, \bar{v} \rangle \leq \|\bar{u}\|^2$ , and the third constraint can be written as  $\langle \bar{u}, \bar{v} \rangle \geq 0$ .

We now describe the approximation algorithm.

*Approximation Algorithm:* We first informally describe the main idea behind the algorithm. The algorithm solves the SDP relaxation and obtains a set of vectors  $\{\bar{u}\}_{u \in V}$ . Now it samples an orthogonal separator, a random set  $S \subset V$ , and returns it. Assume for the moment that  $\alpha = 1$ . Since  $\Pr(v \in S) = \|\bar{v}\|^2$ , we get  $\mathbb{E}[\eta(S)] \geq H$ . The expected size of the cut is at most  $D \times SDP$  by the third property of orthogonal separators; and thus the sparsity is at most

$$\begin{aligned} \min \quad & \frac{1}{w(E)} \sum_{(u,v) \in E} w(u,v) \|\bar{u} - \bar{v}\|^2 \\ \text{subject to:} \quad & \text{for all } u, v, w \in V, \\ & \|\bar{u} - \bar{w}\|^2 + \|\bar{w} - \bar{v}\|^2 \geq \|\bar{u} - \bar{v}\|^2; \\ & \|\bar{u} - \bar{v}\|^2 \geq \|\bar{u}\|^2 - \|\bar{v}\|^2; \\ & \|\bar{u}\|^2 + \|\bar{v}\|^2 \geq \|\bar{u} - \bar{v}\|^2; \\ \text{for all } u \in V, \\ & \sum_{v \in V} \mu(v) \cdot \min\{\|\bar{u} - \bar{v}\|^2, \|\bar{u}\|^2\} \geq (1 - \rho) \|\bar{u}\|^2 \\ & \sum_{v \in V} \|\bar{v}\|^2 \eta(v) \geq H. \end{aligned}$$

Figure 1. SDP Relaxation

$D \times SDP/H \leq 2D \times OPT$ . The second property of orthogonal separators guarantees that if  $\bar{u} \in S$ , then a very small fraction of vectors that are far from  $\bar{u}$  belongs to  $S$  (since the conditional probability  $\Pr(\bar{v} \in S \mid \bar{u} \in S) \leq 1/m$  is very small). And by the spreading constraints, at most  $(1 + \varepsilon)\rho$  fraction of vectors (w.r.t. the measure  $\mu$ ) is close to  $\bar{u}$ . Hence, the total expected measure of  $S$  is at most  $(1 + \varepsilon)\rho + 1/m \leq (1 + 2\varepsilon)\rho$ . We now proceed to the formal argument.

We may assume that  $\varepsilon$  is sufficiently small i.e.,  $\varepsilon \in (0, 1/4)$ . The approximation algorithm guesses approximate value of the weight  $H$ :  $H \leq \eta(S) \leq 2H$ . It sets the length of all vectors  $\bar{u}$  with  $\eta(u) > 2H$  to be 0. Then it solves the SDP and obtains a set of vectors  $X = \{\bar{v}\}_{v \in V}$ . The algorithm finds an orthogonal separator  $S$  with  $m = \max(\varepsilon^{-1} \rho^{-1})$  and  $\beta = \varepsilon$ . For convenience, we let  $S$  be the set of vertices corresponding to vectors belonging to the orthogonal separator rather than the vectors themselves. The algorithm repeats the previous step  $\lceil \alpha^{-1} n^2 \rceil$  times (recall  $\alpha$  is the probabilistic scale of the orthogonal separator) and outputs the best  $S$  satisfying  $0 < \mu(S) < (1 + 10\varepsilon)\rho$ . With an exponentially small probability no  $S$  satisfies this constraint, in which case, the algorithm outputs an arbitrary set satisfying constraints.

*Analysis:* We first estimate the probability of the event “ $u \in S$  and  $\mu(S) < (1 + 10\varepsilon)\rho$ ” for a fixed vertex  $u \in V$ . Let  $A_u = \{v : \|\bar{u} - \bar{v}\|^2 \geq \beta \|\bar{u}\|^2\}$  and  $B_u = \{v : \|\bar{u} - \bar{v}\|^2 < \beta \|\bar{u}\|^2\}$ . We show that only a small fraction of  $A_u$  belongs to  $S$ , and that the set  $B_u$  is small.

From the spreading constraint

$$\sum_{v \in V} \min(\|\bar{u} - \bar{v}\|^2, \|\bar{u}\|^2) \mu(v) \geq (1 - \rho) \|\bar{u}\|^2,$$

and by Markov’s inequality, we get that  $\mu(B_u) \leq \rho/(1 - \beta) \leq (1 + 2\varepsilon)\rho$ . For an arbitrary  $v \in A_u$  (for which  $\bar{v} \neq 0$ ) write  $\|\bar{u} - \bar{v}\|^2 \geq \beta \|\bar{u}\|^2 \geq \beta \min(\|\bar{u}\|^2, \|\bar{v}\|^2)$ . By the

second property of orthogonal separators,  $\Pr(v \in S \mid u \in S) \leq 1/m$ , thus the expected measure  $\mu(A_u \cap S)$  is at most  $\mathbb{E}\mu(A_u \cap S) \leq \varepsilon\rho$ . Now, by the Markov inequality, given that  $u \in S$ , the probability of the bad event “ $\mu(S) \geq (1 + 10\varepsilon)\rho$ ” (and, thus  $\mu(A_u \cap S) \geq 8\varepsilon\rho$ ) is at most  $1/8$ . Each vertex  $u \in V$  belongs to  $S$  with probability  $\alpha\|\bar{u}\|^2$ . Hence,  $u \in S$ , and  $\mu(S) < (1 + 10\varepsilon)\rho$  with probability at least  $3/4$   $\alpha\|\bar{u}\|^2$ .

Finally, we use the third property of orthogonal separators to bound the size of the cut  $\delta(S)$

$$\begin{aligned} \mathbb{E}\delta(S) &= \sum_{(u,v) \in E} |I_S(u) - I_S(v)|w(u,v) \\ &\leq \alpha D \times \sum_{(u,v) \in E} \|\bar{u} - \bar{v}\|^2 w(u,v) \\ &= \alpha D \times SDP \times w(E). \end{aligned}$$

Here, as usual,  $SDP$  denotes the value of the SDP solution; and  $D = O_\varepsilon(\sqrt{\log n \log(1/\delta)})$  is the distortion of  $m$ -orthogonal separators.

Define the function

$$f(S) = \eta(S) - \frac{\delta(S)}{w(E)} \times \frac{H}{4D \times SDP},$$

if  $|S| \neq \emptyset$ ,  $\mu(S) < (1 + 10\varepsilon)\rho$ , and  $f(S) = 0$ , otherwise. The expectation

$$\mathbb{E}f(S) \geq \sum_{u \in V} \frac{3\alpha\|\bar{u}\|^2\eta(u)}{4} - \frac{\alpha H}{4} \geq \frac{\alpha H}{2}.$$

The random variable  $f(S)$  is always bounded by  $2nH$ , thus with probability at least  $\alpha/n$ ,  $f(S) > 0$ . Therefore, with probability exponentially close to 1, after  $\alpha^{-1}n^2$  iterations, the algorithm will find  $S$  with  $f(S) > 0$ . Since  $f(S) > 0$ , we get  $\eta(S) > 0$ ,  $\mu(S) < (1 + 10\varepsilon)\rho$ , and

$$\frac{\delta(S)}{w(E)} \times \frac{1}{\eta(S)} \leq 4D \times \frac{SDP}{H}.$$

This finishes the proof of part I since  $SDP/(2H) \leq \Phi_{\rho, \mu, \eta}(G)$ .

## 2.2. Algorithm II: Small-Set Expansion in General Graphs

We now prove part II of Theorem 8. This algorithm uses an SDP relaxation similar to part I, although we need a few additional constraints. We write a constraint ensuring that “ $\eta(S) \leq 2H$ ” (recall  $H$  is an approximate value of  $\eta(S)$  in the optimal solution): we add spreading constraints for all  $u \in V$ ,

$$\sum_{v \in V} \min\{\|\bar{u} - \bar{v}\|^2, \|\bar{u}\|^2\} \eta(v) \leq 2H\|\bar{u}\|^2,$$

and we let  $m = \max\{\varepsilon^{-1}\rho^{-1}, H^{-1}\rho^{-1}\}$ . We also require

$$\sum_{v \in V} \|v\|^2 \mu(v) \leq \rho. \quad (4)$$

Algorithm II gets  $H$ , the approximate value of the measure  $\eta(S)$ , as input, and thus does not need to guess it.

**Approximation Algorithm.** The algorithm consists of many iterations of a slightly modified Algorithm I. At every step the algorithm obtains a set  $S$  of vertices (returned by Algorithm I) and adds it to the set  $T$ , which is initially empty. Then, the algorithm removes vectors corresponding to  $S$  from the set  $X$ , the SDP solution, and repeats the same procedure till  $\mu(T) \geq \rho/4$  or  $\eta(T) \geq H/4$ . In the end, the algorithm returns the set  $T$  if  $\mu(T) \leq \rho$  and  $\eta(T) \leq H$ , and the last set  $S$  otherwise.

The algorithm changes the SDP solution (by removing some vectors), however we can ignore these changes, since the objective value of the SDP may only decrease and all constraints but (3) are still satisfied. Since the total weight  $\eta(T)$  of removed vertices is at most  $H/4$ , a slightly weaker variant of constraint (3) is satisfied. Namely,

$$\sum_{u \in V} \|\bar{u}\|^2 \eta(u) \geq 3H/4. \quad (3')$$

We now describe the changes in Algorithm I: instead of  $f$ , we define function  $f'$ :

$$f'(S) = \eta(S) - \frac{\delta(S)}{w(E)} \times \frac{H}{4D \times SDP} - \frac{\mu(S)}{4\rho} \times H,$$

if  $|S| \neq \emptyset$ ,  $\mu(S) < (1 + 10\varepsilon)\rho$  and  $\eta(S) \leq (1 + 10\varepsilon)H$  and  $f'(S) = 0$ , otherwise. Notice, that  $f'$  has an extra term comparing to  $f$  and, in order for  $f'(S)$  to be positive, the constraint  $\eta(S) \leq 2(1 + 10\varepsilon)H$  should be satisfied. The new variant of Algorithm I, returns  $S$ , once  $f'(S) > 0$ .

The same argument as before shows that for any given  $u \in V$  conditional on  $u \in S$ ,  $\mu(S) \leq (1 + 10\varepsilon)\rho$  and  $\eta(S) \leq 2(1 + 10\varepsilon)H$  with probability at least  $3/4$ . Then, using a new constraint (4), we get  $\mathbb{E}\mu(S) \leq \alpha\rho$ . Hence, the expectation

$$\mathbb{E}f'(S) \geq \frac{3\alpha \times 3/4 H}{4} - \frac{\alpha H}{4} - \frac{\alpha H}{4} \geq \frac{\alpha H}{16}.$$

Again, after at most  $O(\alpha^{-1}n^2)$  iterations the algorithm will find  $S$  with  $f'(S) > 0$  (and only with exponentially small probability fail)<sup>3</sup>. Then,  $f'(S) > 0$  implies

$$\frac{\delta(S)}{w(E)} \leq 4D \times \frac{SDP}{H} \eta(S); \quad (5)$$

and  $\eta(S) \geq H \times \mu(S)/(4\rho)$ .

The last inequality implies that at every moment  $\eta(T) \geq H \times \mu(T)/(4\rho)$ . Hence, if  $\mu(T) \geq \rho/4$  (recall, this is one of the two conditions, when the algorithm stops), then  $\eta(T) \geq H/16$ . Therefore, if the algorithm returns set  $T$ , then  $\eta(T) \geq H/16$ . If the algorithm returns set  $S$  then either  $\mu(S) \geq 3/4\rho$  and thus  $\eta(S) \geq 3H/16$  or  $\eta(S) \geq 3/4H$ .

Both,  $\mu(T)$  and  $\eta(T)$  are bounded from above by  $\rho$  and  $H$  respectively;  $\mu(S)$  and  $\eta(S)$  are bounded from above by  $(1 + 10\varepsilon)\rho$  and  $2(1 + 10\varepsilon)H$  respectively.

The inequality (5) holds for every set  $S$  added in  $T$ , hence this inequality holds for  $T$ .

<sup>3</sup>In fact, now  $f'(S) \leq 2H$ , thus we need only  $O(\alpha^{-1}n)$  iterations.

### 2.3. Small-Set Expansion in Minor-Closed Graph Families

In this subsection we present Theorem 6. We start by writing an LP relaxation. For every vertex  $u \in V$  we introduce a variable  $x(u)$  taking values in  $[0, 1]$ ; and for every pair of vertices  $u, v \in V$  we introduce a variable  $z(u, v) = z(v, u)$  also taking values in  $[0, 1]$ . In the intended integral solution corresponding to a set  $S \subset V$ ,  $x(u) = 1$  if  $u \in S$ , and  $x(u) = 0$  otherwise;  $z(u, v) = |x(u) - x(v)|$ . (One way of thinking of  $x(u)$  is as the distance to some imaginary vertex  $O$  that never belongs to  $S$ . In the SDP relaxation vertex  $O$  is the origin.) It is instructive to think of  $x(u)$  as an analog of  $\|\bar{u}\|^2$  and of  $z(u, v)$  as an analog of  $\|\bar{u} - \bar{v}\|^2$ .

It is easy to verify that the LP below is a relaxation of the Small-Set Expansion problem. It has a constraint saying that  $z(u, v)$  is a metric (or, strictly speaking, semi-metric). A novelty of the LP is in the third constraint, which is a new spreading constraints for ensuring the size of  $S$  is small.

$$\begin{array}{l}
 \min \quad \sum_{(u,v) \in E} w(u,v)z(u,v) \\
 \text{subject to: for all } u, v, w \in V, \\
 \\
 \quad z(u, v) + z(v, w) \geq z(u, w) \\
 \quad |x(u) - x(v)| \leq z(u, v) \\
 \quad x(u), z(u, v) \in [0, 1] \\
 \\
 \text{for all } u \in V, \\
 \quad \sum_{v \in V} \mu(v) \cdot \min\{x(u), z(u, v)\} \geq (1 - \rho)x(u).
 \end{array}$$

Figure 2. LP Relaxation

We introduce an analog of  $m$ -orthogonal separators for linear programming, which we call LP separators.

**Definition 11** (LP separator). *Let  $G = (V, E)$  be a graph, and let  $\{x(u), z(u, v)\}_{u, v \in V}$  be a set of numbers. We say that a distribution over subsets of  $V$  is an LP separator of  $V$  with distortion  $D \geq 1$ , probability scale  $\alpha > 0$  and separation threshold  $\beta \in (0, 1)$  if the following conditions hold for  $S \subset V$  chosen according to this distribution:*

- For all  $u \in V$ ,  $\Pr(u \in S) = \alpha x(u)$ .
- For all  $u, v \in V$  with  $z(u, v) \geq \beta \min\{x(u), x(v)\}$ ,

$$\Pr(u \in S \text{ and } v \in S) = 0.$$

- For all  $(u, v) \in E$ ,

$$\Pr(I_S(u) \neq I_S(v)) \leq \alpha D \times z(u, v),$$

where  $I_S$  is an indicator for the set  $S$ .

In the full version of the paper, we present an efficient algorithm for an LP separator: given a graph  $G = (V, E)$  excluding  $K_{r,r}$  as a minor, a parameter  $\beta \in (0, 1)$ , and a set of numbers  $\{x(u), z(u, v)\}_{u, v \in V}$  satisfying the triangle inequalities described above (but not necessarily the spreading constraints), the algorithm computes an LP separator with distortion  $O(r^2)$  (for genus  $g$  graphs the distortion is  $O(\log g)$ ). This proves Theorem 6 as follows: by replacing in the algorithms above the SDP relaxation with the LP relaxation, and the orthogonal separators with LP separators, we obtain  $O(r^2)$  approximation algorithm approximation algorithm for SSE in  $K_{r,r}$  excluded-minor graphs. Combined with the framework in Section 3, we consequently obtain an  $O(r^2)$ -approximation algorithm for Min-Max  $k$ -Partitioning and Min-Max-Multiway-Cut on such graphs.

**Theorem 12.** *There exists an algorithm that given a graph  $G = (V, E)$  with an excluded minor  $K_{r,r}$ , a set of numbers  $\{x(u), z(u, v)\}_{u, v \in V}$  satisfying the triangle inequality constraints, and a parameter  $\beta \in [0, 1]$ , returns an LP separator  $S \subset V$  with distortion  $D = r^2 \beta^{-1}$  and separation threshold  $\beta$ .*

### 2.4. From SSE to $\rho$ -Unbalanced Cut

We show that  $\rho$ -Unbalanced Cut and SSE are equivalent, up to some constants, with respect to bicriteria approximation guarantees. Indeed, the two problems are related in the same way that Balanced Cut and Sparsest Cut are. We refer the reader to [16], [20], and omit details from this version of the paper. Our intended application of approximating Min-Max  $k$ -Partitioning (in Section 3), requires a weighted version of the  $\rho$ -Unbalanced Cut problem, as follows.

**Definition 13** (Weighted  $\rho$ -Unbalanced Cut). *The input to this problem is a tuple  $\langle G, y, w, \tau, \rho \rangle$ , where  $G = (V, E)$  is a graph with vertex-weights  $y : V \rightarrow \mathbb{R}_+$ , edge-costs  $w : E \rightarrow \mathbb{R}_{\geq 0}$ , and parameters  $\tau, \rho \in (0, 1]$ . The goal is to find  $S \subseteq V$  of minimum cost  $\delta(S)$  satisfying:*

- 1)  $y(S) \geq \tau \cdot y(V)$ ; and
- 2)  $|S| \leq \rho \cdot n$ .

The unweighted version of the problem (defined in Section 1.2) has  $\tau = \rho$  and unit vertex-weights, i.e.  $y(v) = 1$  for all  $v \in V$ . We focus on the direction of reducing Weighted  $\rho$ -Unbalanced Cut to Weighted Small-Set Expansion, which is needed for our intended application (In the full version of the paper we give reductions in both direction.) Formally, we have the following corollary of Theorem 8. We use  $\text{OPT}_{\langle G, y, w, \tau, \rho \rangle}$  to denote the optimal value of the corresponding weighted  $\rho$ -Unbalanced Cut instance.

**Corollary 14** (Approximating  $\rho$ -Unbalanced Cut). *For every  $\varepsilon > 0$ , there exists a polynomial-time algorithm that given an instance  $\langle G, y, w, \tau, \rho \rangle$  of Weighted  $\rho$ -Unbalanced Cut, finds a set  $S$  satisfying  $|S| \leq \beta \rho n$ ,  $y(S) \geq \tau/\gamma$  and  $\delta(S) \leq$*

$\alpha \cdot \text{OPT}_{\langle G, y, w, \tau, \rho \rangle}$  for  $\alpha = O_\varepsilon(\sqrt{\log n \log(\max(1/\rho, 1/\tau))})$ ,  $\beta = 1 + \varepsilon$  and  $\gamma = O(1)$ .

*Proof:* Let  $S^*$  be an optimal solution to  $\langle G, y, w, \tau, \rho \rangle$ , note that  $|S^*| \leq \rho n$ ,  $y(S^*) \geq \tau \cdot y(V)$  and  $\delta(S^*) = \text{OPT}_{\langle G, y, w, \tau, \rho \rangle}$  the optimal value of this instance. Define two measures on  $V$  as follows. For any  $S \subseteq V$ , set  $\mu(S) := |S|/n$  and  $\eta(S) := y(S)/y(V)$ .

The algorithm guesses  $H \geq \tau$  such that  $H \leq \eta(S^*) \leq 2H$  (see Algorithm I above for an argument why we can guess  $H$ ). Then it invokes the algorithm from part II on  $G$  with measures  $\mu$  and  $\eta$  as defined above, and parameters  $\rho, H$ . The obtained solution  $S$  satisfies  $|S| = \mu(S) \cdot n \leq (1 + \varepsilon)\rho n$  and  $y(S) = \eta(S) \cdot y(V) \geq \Omega_\varepsilon(1) H \cdot y(V) \geq \Omega_\varepsilon(1) \tau \cdot y(V)$ , since  $H \geq \tau$ . Furthermore,  $\delta(S) \leq \alpha \cdot \delta(S^*) \cdot \eta(S)/\eta(S^*) \leq \alpha \cdot \delta(S^*) \cdot \Theta_\varepsilon(1)$ , where  $\alpha = O_\varepsilon(\sqrt{\log n \log(\max(1/\rho, 1/\tau))})$ . ■

### 3. MIN-MAX BALANCED PARTITIONING

In this section, we present our algorithm for Min–Max  $k$ –Partitioning, assuming a subroutine that approximates Weighted  $\rho$ –Unbalanced Cut (which is essentially a rephrasing of Weighted Small-Set Expansion). Our algorithm for Min–Max  $k$ –Partitioning follows follows by a straightforward composition of Theorem 15 and Theorem 17 below. Plugging in for  $(\alpha, \beta, \gamma)$  the values obtained in Section 2 would complete the proof of Theorem 1.

#### 3.1. Uniform Coverings

We first consider a covering relaxation of Min–Max  $k$ –Partitioning and solve it using multiplicative updates. This covering relaxation can alternatively be viewed as a fractional solution to a configuration LP of exponential size.

Let  $\mathcal{C} = \{S \subseteq V : |S| \leq n/k\}$  denote all the vertex-sets that are feasible for a single part. Note that a feasible solution in Min–Max  $k$ –Partitioning corresponds to a partition of  $V$  into  $k$  parts, where each part belongs to  $\mathcal{C}$ . Algorithm 1, described below, *uniformly covers*  $V$  using sets in  $\mathcal{C}$  (actually a slightly larger family than  $\mathcal{C}$ ). It is important to note that its output  $\mathcal{S}$  is a multiset.

**Theorem 15.** *Running Algorithm 1 on an instance of Min–Max  $k$ –Partitioning outputs  $\mathcal{S}$  that satisfies (here  $\text{OPT}$  denotes the optimal value of the instance):*

- 1) For all  $S \in \mathcal{S}$  we have  $\delta(S) \leq \alpha \cdot \text{OPT}$  and  $|S| \leq \beta \cdot n/k$ .
- 2) For all  $v \in V$  we have  $|\{S \in \mathcal{S} : S \ni v\}|/|\mathcal{S}| \geq 1/(5\gamma k)$ .

*Proof:* For an iteration  $t$ , let us denote  $Y^t := \sum_{v \in V} y^t(v)$ . The first assertion of the theorem is immediate from the following claim.

**Claim 16.** *Every iteration  $t$  of Algorithm 1 satisfies  $\delta(S^t) \leq \alpha \cdot \text{OPT}$  and  $|S^t| \leq \beta \cdot n/k$ .*

---

#### Algorithm 1: Covering Procedure

---

```

Set  $t = 1$ , and  $y^1(v) = 1$  for all  $v \in V$ 
while  $\sum_{v \in V} y^t(v) > 1/n$  do
  // Solve the following using algorithm from
  // Corollary 14.
  Let  $S^t \subseteq V$  be the solution for Weighted
   $\rho$ –Unbalanced Cut instance  $\langle G, y^t, w, \frac{1}{k}, \frac{1}{k} \rangle$ .
  Set  $\mathcal{S} = \mathcal{S} \cup \{S^t\}$ .
  // Update the weights of the covered vertices.
  for every  $v \in V$  do
    [ Set  $y^{t+1}(v) = \frac{1}{2} \cdot y^t(v)$  if  $v \in S^t$ , and
    [  $y^{t+1}(v) = y^t(v)$  otherwise.
  Set  $t = t + 1$ .
return  $\mathcal{S}$ 

```

---

*Proof:* It suffices to show that the optimal value of the Weighted  $\rho$ –Unbalanced Cut instance  $\langle G, y^t, w, \frac{1}{k}, \frac{1}{k} \rangle$  is at most  $\text{OPT}$ . To see this, consider the optimal solution  $\{S_i^*\}_{i=1}^k$  of the original Min–Max  $k$ –Partitioning instance. We have  $|S_i^*| \leq n/k$  and  $w(\delta(S_i^*)) \leq \text{OPT}$  for all  $i \in [k]$ . Since  $\{S_i^*\}_{i=1}^k$  partitions  $V$ , there is some  $j \in [k]$  with  $y^t(S_j^*) \geq Y^t/k$ . It now follows that  $S_j^*$  is a feasible solution to the Weighted  $\rho$ –Unbalanced Cut instance  $\langle G, y^t, w, \frac{1}{k}, \frac{1}{k} \rangle$ , with objective value at most  $\text{OPT}$ , which proves the claim. ■

We proceed to prove the second assertion of Theorem 15. Let  $\ell$  denote the number of iterations of the while loop, for the given Min–Max  $k$ –Partitioning instance. For any  $v \in V$ , let  $N_v$  denote the number of iterations  $t$  with  $S^t \ni v$ . Then, by the  $y$ –updates we have  $y^{\ell+1}(v) = 1/2^{N_v}$ . Moreover, the termination condition implies that  $y^{\ell+1}(v) \leq 1/n$  (since  $Y^{\ell+1} \leq 1/n$ ). Thus we obtain  $N_v \geq \log_2 n$  for all  $v \in V$ . From the approximation guarantee of the Weighted  $\rho$ –Unbalanced Cut algorithm, it follows that  $y^t(S^t) \geq \frac{1}{\gamma k} \cdot Y^t$  in every iteration  $t$ . Thus  $Y^{\ell+1} = Y^\ell - \frac{1}{2} \cdot y^\ell(S^\ell) \leq \left(1 - \frac{1}{2\gamma k}\right) \cdot Y^\ell$ . This implies that

$$Y^\ell \leq \left(1 - \frac{1}{2\gamma k}\right)^{\ell-1} \cdot Y^1 = \left(1 - \frac{1}{2\gamma k}\right)^{\ell-1} \cdot n.$$

However  $Y^\ell > 1/n$  since the algorithm performs  $\ell$  iterations. Thus,  $\ell \leq 1 + 4\gamma k \cdot \ln n \leq 5\gamma k \cdot \log_2 n$ . This proves  $|\{S \in \mathcal{S} : S \ni v\}|/|\mathcal{S}| = N_v/\ell \geq (5\gamma)^{-1} k^{-1}$ . ■

#### 3.2. Aggregation

The aggregation process, which might be of independent interest, transforms a cover of  $G$  into a partition. Intuitively, we first let the sets randomly compete with each other over the vertices so as to form a partition; then, to make sure no set has large cost, we repeatedly fix the partition locally, and use a potential function to track progress.



**Theorem 17.** *Algorithm 2 is a randomized polynomial-time algorithm that when given a graph  $G = (V, E)$ , an  $\varepsilon \in (0, 1)$ , and a cover  $\mathcal{S}$  of  $V$  that satisfies: (i) every vertex in  $V$  is covered by at least  $c/k$  fraction of sets  $S \in \mathcal{S}$ , for  $c \in (0, 1]$ ; and (ii) all  $S \in \mathcal{S}$  satisfy  $|S| \leq 2n/k$  and  $\delta(S) \leq B$ ; the algorithm outputs a partition  $\mathcal{P}$  of  $V$  into at most  $k$  sets such that for all  $P \in \mathcal{P}$  we have  $|P| \leq 2(1 + \varepsilon)n/k$  and  $\mathbb{E}[\max \delta(P) : P \in \mathcal{P}] \leq 8B/(c\varepsilon)$ .*

---

**Algorithm 2:** Aggregation Procedure

---

1 *Sampling*

Sort sets in  $\mathcal{S}$  in a random order:  $S_1, S_2, \dots, S_{|\mathcal{S}|}$ .

Let  $P_i = S_i \setminus \cup_{j < i} S_j$ .

2 *Replacing Expanding Sets with Sets from  $\mathcal{S}$*

**while** there is a set  $P_i$  such that  $\delta(P_i) > 2B$  **do**

└ Set  $P_i = S_i$ , and for all  $j \neq i$ , set  $P_j = P_j \setminus S_i$ .

3 *Aggregating*

Let  $B' = \max\{\frac{1}{k} \sum_i \delta(P_i), 2B\}$ .

**while** there are  $P_i \neq \emptyset, P_j \neq \emptyset$  ( $i \neq j$ ) such that

$|P_i| + |P_j| \leq 2(1 + \varepsilon)n/k$  and

$\delta(P_i) + \delta(P_j) \leq 2B'\varepsilon^{-1}$  **do**

└ Set  $P_i = P_i \cup P_j$  and set  $P_j = \emptyset$ .

4 **return** all non-empty sets  $P_i$ .

---

**Analysis:** 1. Observe that after step 1 the collection of sets  $\{P_i\}$  is a partition of  $V$  and  $P_i \subset S_i$  for every  $i$ . Particularly,  $|P_i| \leq |S_i| \leq 2n/k$ . Note, however, that the bound  $\delta(P_i) \leq B$  may be violated for some  $i$ . We now prove that  $\mathbb{E}[\sum_i \delta(P_i)] \leq 2kB/c$ .

Fix an  $i \leq |\mathcal{S}|$  and estimate the expected weight of edges  $E(P_i, \cup_{j > i} P_j)$  given that  $S_i = S$ . If an edge  $(u, v)$  belongs to  $E(P_i, \cup_{j > i} P_j)$  then  $(u, v) \in E(S_i, V \setminus S_i) = E(S, V \setminus S)$  and both  $u, v \notin \cup_{j < i} S_j$ . For any edge  $(u, v) \in \delta(S)$  (with  $u \in S, v \notin S$ ),

$$\begin{aligned} \Pr((u, v) \in E(P_i, \cup_{j > i} P_j) \mid S_i = S) \\ \leq \Pr(v \notin \cup_{j < i} S_j \mid S_i = S) \leq (1 - c/k)^{i-1}, \end{aligned}$$

since  $v$  is covered by at least  $c/k$  fraction of sets in  $\mathcal{S}$  and is not covered by  $S_i = S$ . Hence,

$$\begin{aligned} \mathbb{E}[w(E(P_i, \cup_{j > i} P_j)) \mid S_i = S] \\ \leq (1 - c/k)^{i-1} \delta(S) \leq (1 - c/k)^{i-1} B, \end{aligned}$$

and  $\mathbb{E}[w(E(P_i, \cup_{j > i} P_j))] \leq (1 - c/k)^{i-1} B$ . Therefore, the total expected weight of edges crossing the boundary of  $P_i$ 's is at most  $\sum_{i=0}^{\infty} (1 - c/k)^i B = kB/c$ , and  $\mathbb{E}[\sum_i \delta(P_i)] \leq 2kB/c$ .

2. After each iteration of step 2, the following invariant holds: the collection of sets  $\{P_i\}$  is a partition of  $V$  and  $P_i \subset S_i$  for all  $i$ . Particularly,  $|P_i| \leq |S_i| \leq 2n/k$ . The key observation is that at every iteration of the ‘‘while’’ loop, the

sum  $\sum_j \delta(P_j)$  decreases by at least  $2B$ . This is due to the following uncrossing argument:

$$\begin{aligned} \sum_{j \neq i} \delta(P_j \setminus S_i) &\leq \\ &\leq \sum_{j \neq i} \left( \delta(P_j) + w(E(P_j \setminus S_i, S_i)) - w(E(S_i \setminus P_j, P_j)) \right) \\ &\leq \left( \sum_{j \neq i} \delta(P_j) \right) + \underbrace{w(E(V \setminus S_i, S_i))}_{\delta(S_i)} - \underbrace{w(E(P_i, V \setminus P_i))}_{\delta(P_i)} \\ &= \left( \sum_j \delta(P_j) \right) + \delta(S_i) - \delta(P_i) \\ &\leq \left( \sum_{j \neq i} \delta(P_j) \right) - B. \end{aligned}$$

We used that  $P_i \subset S_i$ , all  $P_j$  are disjoint,  $\cup_{j \neq i} (P_j \setminus S_i) \subset V \setminus S_i$ ,  $P_i \subset S_i \setminus P_j$ ,  $\cup_{j \neq i} P_j = V \setminus P_i$ . Then,

$$\begin{aligned} \delta(S_i) + \sum_{j \neq i} \delta(P_j \setminus S_i) &\leq \left( \sum_j \delta(P_j) \right) - B \\ &+ (\delta(S_i) - \delta(P_i)) \leq \left( \sum_j \delta(P_j) \right) - 2B. \end{aligned}$$

Hence, the number of iterations of the loop in step 2 is always polynomially bounded and after the last iteration  $\mathbb{E}[\sum_i \delta(P_i)] \leq 2kB/c$  (the expectation is over random choices at step 1; the step 2 does not use random bits). Hence,  $\mathbb{E}[B'] \leq 4B/c$ .

3. The following analysis holds conditional on any value of  $B'$ . After each iteration of step 3, the following invariant holds: the collection of sets  $\{P_i\}$  is a partition of  $V$ . Moreover,  $|P_i| \leq 2(1 + \varepsilon)n/k$  and  $\delta(P_i) \leq 2B'\varepsilon^{-1}$  (note: after step 2,  $\delta(P_i) \leq 2B \leq B'$  for each  $i$ ).

When the loop terminates, we obtain a partition of  $V$  into sets  $P_i$  satisfying  $|P_i| \leq 2(1 + \varepsilon)n/k$ ,  $\sum_i |P_i| = n$ ,  $\delta(P_i) \leq 2B'\varepsilon^{-1}$ ,  $\sum_i \delta(P_i) \leq kB'$ , such that no two sets can be merged without violating above constraints. Hence by Lemma 18 below (with  $a_i = |P_i|$  and  $b_i = \delta(P_i)$ ), the number of non-empty sets is at most

$$2 \frac{n}{2(1 + \varepsilon)n/k} + \frac{kB'}{2B'\varepsilon^{-1}} = (1 + \varepsilon)^{-1}k + (\varepsilon/2)k \leq k. \quad \blacksquare$$

**Lemma 18 (Greedy Aggregation).** *Let  $a_1, \dots, a_t$  and  $b_1, \dots, b_t$  be two sequences of nonnegative numbers satisfying the following constraints  $a_i < A$ ,  $b_i < B$ ,  $\sum_{i=1}^t a_i \leq S$  and  $\sum_{i=1}^t b_i \leq T$  (for some positive real numbers  $A, B, S$ , and  $T$ ). Moreover, assume that for every  $i$  and  $j$  ( $i \neq j$ ) either  $a_i + a_j > A$  or  $b_i + b_j > B$ . Then,  $t < S/A + T/B + \max(S/A, T/B, 1)$ .*

*Proof:* By rescaling we assume that  $A = 1$  and  $B = 1$ . Moreover, we may assume that  $\sum_{i=1}^t a_i < S$  and  $\sum_{i=1}^t b_i < T$  by slightly decreasing values of all  $a_i$  and  $b_i$  so that all inequalities still hold.

We write two linear programs. The first LP ( $LP_I$ ) has variables  $x_i$  and constraints  $x_i + x_j \geq 1$  for all  $i, j$  such that  $a_i + a_j \geq 1$ . The second LP ( $LP_{II}$ ) has variables  $y_i$  and constraints  $y_i + y_j \geq 1$  all  $i, j$  such that  $b_i + b_j \geq 1$ . The LP objectives are to minimize  $\sum_i x_i$  and to minimize  $\sum_i y_i$ . Note, that  $\{a_i\}$  is a feasible point for  $LP_I$  and  $\{b_i\}$  is a feasible point for  $LP_{II}$ . Thus, the optimum values of  $LP_I$  and  $LP_{II}$  are strictly less than  $S$  and  $T$  respectively.

Observe that both LPs are half-integral. Consider optimal solutions  $x_i^*, y_j^*$  where  $x_i^*, y_j^* \in \{0, 1/2, 1\}$ . Note that for every  $i, j$  either  $x_i^* + x_j^* \geq 1$  or  $y_i^* + y_j^* \geq 1$ . Consider several cases. If for all  $i$ ,  $x_i^* + y_i^* \geq 1$ , then  $t < S + T$ , since  $\sum_{i=1}^t (x_i^* + y_i^*) < S + T$ . If for some  $j$ ,  $x_j^* + y_j^* = 0$  (and hence  $x_j^* = y_j^* = 0$ ), then  $x_i^* + y_i^* \geq 1$  for  $i \neq j$  and, thus,  $t < S + T + 1$ . Finally, assume that for some  $j$ ,  $x_j^* + y_j^* = 1/2$ , and w.l.o.g.  $x_j^* = 1/2$  and  $y_j^* = 0$ . The number of  $i$ 's with  $x_i^* \neq 0$  is (strictly) bounded by  $2S$ . For the remaining  $i$ 's,  $x_i^* = 0$  and hence  $y_i^* = 1$  (because  $y_i^* = x_i^* + y_j^* \geq 1$ ), and thus the number of such  $i$ 's is (strictly) bounded by  $T$ . ■

#### REFERENCES

- [1] K. Andreev and H. Racke, "Balanced graph partitioning," *Theor. Comp. Sys.*, vol. 39, no. 6, pp. 929–939, 2006.
- [2] S. Arora, S. Rao, and U. V. Vazirani, "Geometry, flows, and graph-partitioning algorithms," *Commun. ACM*, vol. 51, no. 10, pp. 96–105, 2008.
- [3] Y. Aumann and Y. Rabani, "An  $O(\log k)$  approximate min-cut max-flow theorem and approximation algorithm," *SIAM J. Comput.*, vol. 27, no. 1, 1998.
- [4] N. Bansal, K.-W. Lee, V. Nagarajan, and M. Zafer, "Minimum congestion mapping in a cloud," in *PODC*, 2011.
- [5] G. Calinescu, H. J. Karloff, and Y. Rabani, "An improved approximation algorithm for multiway cut," *J. Comput. Syst. Sci.*, vol. 60, no. 3, pp. 564–574, 2000.
- [6] J. Cheeger, B. Kleiner, and A. Naor, "A  $(\log n)^{\Omega(1)}$  integrality gap for the sparsest cut SDP," in *FOCS*. IEEE Computer Society, 2009, pp. 555–564.
- [7] E. Chlamtac, K. Makarychev, and Y. Makarychev, "How to play unique games using embeddings," in *FOCS*, 2006, pp. 687–696.
- [8] N. M. M. K. Chowdhury, M. R. Rahman, and R. Boutaba, "Virtual network embedding with coordinated node and link mapping," in *INFOCOM*. IEEE, 2009, pp. 783–791.
- [9] G. Even, J. Naor, S. Rao, and B. Schieber, "Fast approximate graph partitioning algorithms," *SIAM J. Comput.*, vol. 28, no. 6, pp. 2187–2214, 1999.
- [10] U. Feige, R. Krauthgamer, and K. Nissim, "On cutting a few vertices from a graph," *Discrete Appl. Math.*, vol. 127, pp. 643–649, May 2003.
- [11] D. R. Karger, P. Klein, C. Stein, M. Thorup, and N. E. Young, "Rounding algorithms for a geometric embedding of minimum multiway cut," *Math. Oper. Res.*, vol. 29, no. 3, pp. 436–461, 2004.
- [12] P. Klein, S. A. Plotkin, and S. Rao, "Excluded minors, network decomposition, and multicommodity flow," in *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, ser. STOC. New York, NY, USA: ACM, 1993, pp. 682–690.
- [13] R. Krauthgamer, J. Naor, and R. Schwartz, "Partitioning graphs into balanced components," in *SODA*, 2009, pp. 942–949.
- [14] J. R. Lee and A. Naor, " $L_p$  metrics on the Heisenberg group and the Goemans-Linial conjecture," in *FOCS*. IEEE Computer Society, 2006, pp. 99–108.
- [15] J. R. Lee and A. Sidiropoulos, "Genus and the geometry of the cut graph," in *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2010, pp. 193–201.
- [16] T. Leighton and S. Rao, "Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms," *J. ACM*, vol. 46, no. 6, pp. 787–832, 1999.
- [17] N. Linial, E. London, and Y. Rabinovich, "The geometry of graphs and some of its algorithmic applications," *Combinatorica*, vol. 15, no. 2, 1995.
- [18] H. Räcke, "Optimal hierarchical decompositions for congestion minimization in networks," in *STOC*, 2008, pp. 255–264.
- [19] P. Raghavendra and D. Steurer, "Graph expansion and the unique games conjecture," in *STOC*, 2010, pp. 755–764.
- [20] P. Raghavendra, D. Steurer, and P. Tetali, "Approximations for the isoperimetric and spectral profile of graphs and related parameters," in *STOC*, 2010, pp. 631–640.
- [21] P. Raghavendra, D. Steurer, and M. Tulsiani, "Reductions between expansion problems," 2010, manuscript.
- [22] Z. Svitkina and É. Tardos, "Min-max multiway cut," in *APPROX-RANDOM*, 2004, pp. 207–218.
- [23] M. Yu, Y. Yi, J. Rexford, and M. Chiang, "Rethinking virtual network embedding: substrate support for path splitting and migration," *ACM SIGCOMM CCR*, vol. 38(2), pp. 17–29, 2008.
- [24] Y. Zhu and M. H. Ammar, "Algorithms for assigning substrate network resources to virtual network components," in *INFOCOM*. IEEE, 2006.